鲁棒的交叉熵模糊聚类算法 *

姚　兰，严寒冰，蔚泽峰

(成都信息工程大学 控制工程学院，成都 610225)

摘　要：针对模糊 C-均值聚类算法对噪声敏感、容易收敛到局部极小值的问题，提出一种基于交叉熵的模糊聚类算法。该算法通过引入交叉熵重新定义了传统 FCM 算法的目标函数，利用交叉熵度量样本隶属度之间的差异性，并采用拉格朗日求解方法和朗伯 W 函数解决了目标函数的优化问题，此外，分析了样本划分矩阵的分布情况，依据分布特性对噪声样本进行识别。人工数据集合和标准数据集加噪的实验结果表明，该算法提高了传统 FCM 算法的抗干扰能力，具有更强的鲁棒性，噪声样本识别的准确率较高。

关键词：模糊聚类；交叉熵；模糊 C-均值聚类；聚类性能

中图分类号：TP301.6　　doi: 10.3969/j.issn.1001-3695.2018.03.0198

Robust fuzzy clustering algorithm based on cross entropy

Yao Lan, Yan Hanbing, Wei Zefeng

(*School of Control Engineering*, *University of Information Technology*, *Chengdu* 610225, *China*)

Abstract: For the problem that the traditional fuzzy C-means clustering algorithm is easy to be affected by noise data, this paper proposed a fuzzy clustering algorithm based on the cross entropy. It introduced the cross entropy to the objective function of FCM algorithm to measure the difference between membership function of data, and use Lagrange method and Lambert W function to solve the optimization problem of the objective function. The algorithm could identify the noise samples according to the characteristics of the sample partition matrix. The experiment results of a synthetic data set and a standard data set with noisy show that the algorithm is more robust and has better clustering results.

Key words: fuzzy clustering; cross entropy; FCM; clustering performance

0　引言

自 1969 年 Ruspini[1]首先将模糊理论引入聚类分析，基于模糊理论的聚类分析方法就受到了研究人员的广泛关注，提出了多种模糊聚类分析方法。其中，模糊 C-均值聚类算法（fuzzy C-means，FCM）[2]因其具有设计简单、解决问题范围广且易于实现的特点，已成为模糊聚类算法中一种经典方法在诸多领域得到了广泛应用。但理论研究与实验表明，FCM 算法存在一些固有缺陷，如对噪声数据极其敏感，容易收敛到局部极小点等。为此，研究人员通过修改聚类算法的目标函数和弱化隶属度的约束条件提出了很多算法。代表性算法主要有可能性聚类算法[3]（possiblistic C-means，PCM）及改进的 IPCM 算法[4]、可能性 FCM 算法（possiblistic fuzzy C-means,PFCM）[5,6]、引入样本加权策略的 WFCM 算法（weighted fuzzy C-means ,WFCM）[7]及相关算法[8,9]、基于噪声模型的聚类算法[10]及增强模糊划分的聚类算法[11,12]。这些算法在一定程度上改进了 FCM 所面临的问题，使其在实际应用中有更强的适用能力。

受信息论中采用熵度量系统信息量大小的思想启发，学者将熵引入 FCM 算法展开了基于熵的模糊聚类算法研究。这类方法使得样本点隶属度求解公式具有了高斯分布特性，从而表现出更强的抗噪性。文献[13]采用熵函数作为目标函数的一个组成部分，提出了极大熵聚类算法 MEC，但该方法对例外点较敏感，例外点的干扰常使得到的聚类中心严重偏离。文献[14]提出了鲁棒的极大熵聚类算法 (Robust maximum entropy clustering,RMEC)，提高了算法对例外点的抗干扰能力。在引入模糊熵的基础上，文献[15]通过结合犹豫度重新定义目标函数提出了直觉模糊 C-均值聚类算法，该算法对噪声数据抑制具有一定的效果，但效果不明显，算法复杂性很高。文献[16]采用相对熵替代熵提出了相对熵模糊 C-均值聚类算法(Relative entropy fuzzy c-means clustering,REFCM)。文献[17]提出了基于广义熵的模糊聚类算法，并利用增广拉格朗日法和 Hopfield 神经网络求解。文献[18] 提出了基于广义熵的可能性模糊 C-均值聚类算法，能够更加正确地获得含噪声数据的聚类中心。

考虑到熵只针对单个变量，不能有效度量两个随机分布的

收稿日期：2018-03-21；**修回日期：**2018-05-04　**基金项目：**四川省教育厅重点项目（17ZA0073）；成都信息工程大学引进人才项目（KYTZ201522）

作者简介：姚兰（1980-），女，副教授，博士，主要研究方向为智能信息处理、模式识别与图像处理（dancyyao@cuit.edu.cn）；严寒冰（1977-），男，副教授，博士，主要研究方向为智能信息处理；蔚泽峰（1976-），男，副教授，硕士，主要研究方向为智能信息处理.

差异性以及不满足对称性等缺陷，本文将交叉熵引入传统 FCM 算法的目标函数，提出一种交叉熵模糊聚类算法(cross entropy fuzzy C-means clustering,CEFCM)，采用交叉熵有效度量各类样本隶属度之间的差异性，使目标函数通过优化所获得的划分矩阵中每个样本隶属度不仅受距离影响，而且还受制于交叉熵。

1　模糊交叉熵

基于 Kullback[19]在信息论中对交叉熵的定义，对于论域 $X=\{x_1,x_2,\cdots,x_n\}$ 的两个模糊集 A 和 B，A相对于B的模糊交叉熵的对称形式定义为

$$D(A,B)=\sum_{i=1}^{N}\mu_A(x_i)\ln\frac{\mu_A(x_i)}{\mu_B(x_i)}+\sum_{i=1}^{N}\mu_B(x_i)\ln\frac{\mu_B(x_i)}{\mu_A(x_i)} \tag{1}$$

其中：$\mu_A(x_i)$ 和 $\mu_B(x_i)$ 分别表示X中元素 x_i 分别属于 A 或 B 的隶属度。

交叉熵是下凸函数，当 A 或 B 完全一致时可取得最小零值。

尽管交叉熵不是真正意义上的几何距离，但是用来度量模糊集之间的差异性十分有效。

2　CEFCM 聚类算法

2.1　算法的基本思想

聚类就是按照数据对象的差异性将一个数据集进行区分和分类的过程，差异性越大越有利于削弱无关信息、保留相关信息，从而形成良好的聚类结果。据此，本文将度量模糊集之间差异性的模糊交叉熵引入 FCM 的目标函数，提出了 CEFCM 算法，构造的目标函数为

$$\min J(U,V,c)=\sum_{i=1}^{c}\sum_{j=1}^{N}\mu_{ij}d_{ij}^2\text{-}\theta\left(\sum_{j=1}^{N}\sum_{i=1}^{c}\sum_{\substack{k=1\\k\neq i}}^{c}\mu_{ij}\ln\left(\frac{\mu_{ij}}{\mu_{kj}}\right)+\mu_{kj}\ln\left(\frac{\mu_{kj}}{\mu_{ij}}\right)\right), \tag{2}$$

$$\text{s.t.}\ \sum_{i=1}^{c}\mu_{ij}=1,\forall j;0<\sum_{j=1}^{N}\mu_{ij},\forall i;\quad \mu_{ij}\in[0,1],\forall i,j. \tag{3}$$

其中：d_{ij}^2 是第 j 个样本到第 i个聚类中心的欧式距离，μ_{ij} 是第j个样本对第i个聚类中心的隶属度，c是聚类中心数目，N是样本个数，θ 是交叉熵的调整系数，决定了交叉熵的影响程度。

该目标函数包括两项，第一项为原 FCM 目标函数，第二项为交叉熵函数，本质上度量各个样本隶属度之间的差异程度。通过将两项进行融合得到的 CEFCM 算法，其实质上是使得目标函数优化时所获得的划分矩阵中各个样本的隶属度不仅受距离影响而且还受制于交叉熵，迭代过程中不容易陷入局部最小值。从目标函数的构成显然可以看出，第一项取值最小、第二项取值最大时，目标函数可达到最小值。这表明所有样本与其聚类中心之间的距离最小，类内数据比较集中，而且各类样本隶属度之间的差异最大，类间划分清晰明了，形成了最优的划分结果。

此外，理论上最终所获得的划分矩阵中，数据样本和噪声样本的隶属度分布互不相同：数据样本的各类隶属度值中必然某一类的隶属度值较大、其余类的隶属度值较小，而噪声样本对各类的隶属度值均较小。为此，依据数据样本和噪声样本隶属度分布特点的差异性，则可从样本集中筛选隶属度值均较低的样本作为噪声样本，完成噪声样本的识别。

2.2　算法描述

对于上节给出的目标函数，本节将给出具体推导和算法步骤。

利用拉格朗日乘数 $\lambda_j(j=1,2,\cdots,N)$，构造目标优化函数为

$$J=\sum_{i=1}^{c}\sum_{j=1}^{N}\mu_{ij}d_{ij}^2\text{-}\theta\left(\sum_{j=1}^{N}\sum_{i=1}^{c}\sum_{\substack{k=1\\k\neq i}}^{c}\left(\mu_{ij}\ln\left(\frac{\mu_{ij}}{\mu_{kj}}\right)+\mu_{kj}\ln\left(\frac{\mu_{kj}}{\mu_{ij}}\right)\right)\right)-\sum_{j=1}^{N}\lambda_j\left(\sum_{i=1}^{c}\mu_{ij}-1\right) \tag{4}$$

其中样本对其聚类中心的隶属度值 μ_{ij} 是相互独立的。设

$$\begin{aligned}L&=\mu_{ij}d_{ij}^2-\theta\left(\sum_{\substack{k=1\\k\neq i}}^{c}\left(\mu_{ij}\ln\left(\frac{\mu_{ij}}{\mu_{kj}}\right)+\mu_{kj}\ln\left(\frac{\mu_{kj}}{\mu_{ij}}\right)\right)\right)-\lambda_j(\mu_{ij}-1)\\&=\mu_{ij}d_{ij}^2\text{-}\theta\left(\mu_{ij}\ln(\mu_{ij})-\mu_{ij}\sum_{\substack{k=1\\k\neq i}}^{c}\ln(\mu_{kj})+\sum_{\substack{k=1\\k\neq i}}^{c}\mu_{kj}\ln(\mu_{kj})-\ln(\mu_{ij})\sum_{\substack{k=1\\k\neq i}}^{c}\mu_{kj}\right)-\lambda_j(\mu_{ij}-1)\end{aligned} \tag{5}$$

显然，目标函数 J 关于 U 的极小值求解问题可分解为 L 关于各个隶属函数 μ_{ij} 的极小值求解问题，即目标函数最小化的第一个必要条件 $\frac{\partial J}{\partial U}=0$ 等价于 $\frac{\partial L}{\partial \mu_{ij}}=0$。

对 μ_{ij} 求偏导 $\frac{\partial L}{\partial \mu_{ij}}=0$，可得

$$d_{ij}^2-\theta\left(\ln\mu_{ij}+1-\sum_{\substack{k=1\\k\neq i}}^{c}\ln(\mu_{kj})-\frac{\sum_{k=1,k\neq i}^{c}\mu_{kj}}{\mu_{ij}}\right)-\lambda_j=0 \tag{6}$$

式(6)无法进行直接求解，考虑取值范围的一致性，本文采用$\exp(-y_{ij})$和$\exp(-y_{kj})$替代μ_{ij}和μ_{kj}，则式(6)改写为

$$d_{ij}^2+\theta y_{ij}-\theta-\theta\sum_{\substack{k=1\\k\neq i}}^{c}y_{kj}+ \theta\exp(y_{ij})\sum_{\substack{k=1\\k\neq i}}^{c}\exp(-y_{kj})-\lambda_j=0 \tag{7}$$

令 $d_{ij}^2-\theta-\theta\sum_{\substack{k=1\\k\neq i}}^{c}y_{kj}-\lambda_j=K$，则式（7）改写为

$$K+\theta y_{ij}+\theta\exp(y_{ij})\sum_{\substack{k=1\\k\neq i}}^{c}\exp(-y_{kj})=0 \tag{8}$$

对式（8）求解 y_{ij} 可得

$$y_{ij}=-\frac{K}{\theta}-W_0\left(exp\left(-\frac{K}{\theta}\right)\sum_{\substack{k=1\\k\neq i}}^{c}\exp(-y_{kj})\right) \tag{9}$$

从而

$$\mu_{ij}=\exp\left(\frac{K}{\theta}+W_0\left(exp\left(-\frac{K}{\theta}\right)\sum_{\substack{k=1\\k\neq i}}^{c}\mu_{kj}\right)\right)=\frac{\sum_{k=1,k\neq i}^{c}\mu_{kj}}{W_0\left(exp\left(-\frac{K}{\theta}\right)\sum_{k=1,k\neq i}^{c}\mu_{kj}\right)} \tag{10}$$

其中 $W_0(\bullet)$ 是朗伯 W 函数[20]，满足 $W(Z)\exp(W(Z))=Z$.

对于聚类中心 v_i，一般方法是通过求解 $\frac{\partial L}{\partial v_i}=0$ 得到迭代更新表达式。考虑本文中 d_{ij}^2 表示欧拉距离，则 v_i 采用如下的更新表达式:

$$v_i=\frac{\sum_{j=1}^{N}\mu_{ij}x_j}{\sum_{j=1}^{N}\mu_{ij}} \tag{11}$$

2.3　参数 λ_j 的选取

利用约束条件 $\sum_{i=1}^{c}\mu_{ij}=1,\ \forall j$，可得

$$\sum_{i=1}^{c}\frac{\sum_{k=1,k\neq i}^{c}\ln(\mu_{kj})}{W_0\left(exp\left(-\frac{K}{\theta}\right)\sum_{k=1,k\neq i}^{c}\ln(\mu_{kj})\right)}=1 \tag{12}$$

观察式（12）可发现，λ_j 和隶属函数相互依赖，无法直接求解 λ_j，并且利用 $\frac{\partial J}{\partial\lambda_j}=0$ 也无法求解。为此，本文利用μ_{ij}的条件来确定 λ_j 的取值范围：

1） $\mu_{ij}\geq 0,\ \forall i,j$

该条件要求式（10）是非负的，在该式中 $\sum_{\substack{k=1\\k\neq i}}^{c}\mu_{kj}\geq 0$，所以 $W_0\left(exp\left(-\frac{K}{\theta}\right)\sum_{\substack{k=1\\k\neq i}}^{c}\mu_{kj}\right)$ 必须是非负的。对于函数 $W_0(\bullet)$，$W_0(0)=0$，当 $Z>0$时，$W_0(\bullet)>0$。本文中 $Z=exp\left(-\frac{K}{\theta}\right)\sum_{\substack{k=1\\k\neq i}}^{c}\mu_{kj}>0$，于是 $\mu_{ij}\geq 0$，μ_{ij} 的取值和 λ_j 无关。

2） $\mu_{ij}\leq 1,\ \forall i,j$

该条件要求各个样本对于聚类中心的隶属度必须小于 1，即

$$\sum_{\substack{k=1\\k\neq i}}^{c}\mu_{kj}\leq W_0\left(exp\left(-\frac{K}{\theta}\right)\sum_{\substack{k=1\\k\neq i}}^{c}\mu_{kj}\right) \tag{13}$$

从而得到

$$\lambda_j\geq d_{ij}^2-\theta+\theta\sum_{\substack{k=1\\k\neq i}}^{c}\mu_{kj}+\theta\sum_{\substack{k=1\\k\neq i}}^{c}\ln(\mu_{kj}) \tag{14}$$

于是，根据式(14) 可以确定每次迭代过程中λ_j的取值。

2.4　算法步骤

基于上述讨论，本文 CEFCM 算法的计算步骤如下：

a)确定聚类数目 $c(2\leq c\leq N)$ 和交叉熵系数 θ，以及迭代阈值 ε。

b)初步确定隶属度$\mu_{ij}^{(0)}$和初始的聚类中心$v_i^{(0)}$，$t=1$。

c)计算样本和各个聚类中心之间的距离 d_{ij}^2。

d)按式（14）确定拉格朗日乘子 λ_j。

e)按式（10）更新隶属度 $\mu_{ij}^{(t)}$。

f)按式（11）更新聚类中心$v_i^{(t)}$。

g)如果 $\left\|V^{(t-1)}-V^{(t)}\right\|<\varepsilon$，则算法停止；否则 $t=t+1$，并转步骤 c)。

2.5　算法时间复杂度分析

经典 FCM 算法的时间复杂度为 $O(Nc^2p)$，其中 N 代表数据集的大小，c 为类别数，p 为数据集的维度。对于本文 CEFCM 算法，由于在目标函数中引入了交叉熵，其时间复杂度为

$O(2Nc\log(Nc))$，故整体时间复杂度为 $O(Nc^2p+2Nc\log(Nc))$。虽然本文 CEFCM 算法增大了时间开销，但其聚类效果显著提高。

3　实验结果与分析

为了验证本文 CEFCM 算法的性能，本节使用两组不同的数据集进行实验。第 1 组实验使用人工数据集测试 CEFCM 算法的性能，第 2 组实验使用 IRIS 标准数据集对 CEFCM 算法进行分析与评估。

3.1　人工数据集实验

实验中，本文采用聚类有效性指数对比算法性能。鉴于文献[21]指出基于隶属度矩阵的有效性指数更加适用于实际应用中样本数据含有噪声的情况，本文基于隶属度矩阵计算紧致度和离散度。紧致度定义如下：

$$C=\sum_{i=1}^{c}\sum_{j=1}^{n}C_{ij}\,, \tag{15}$$

其中:

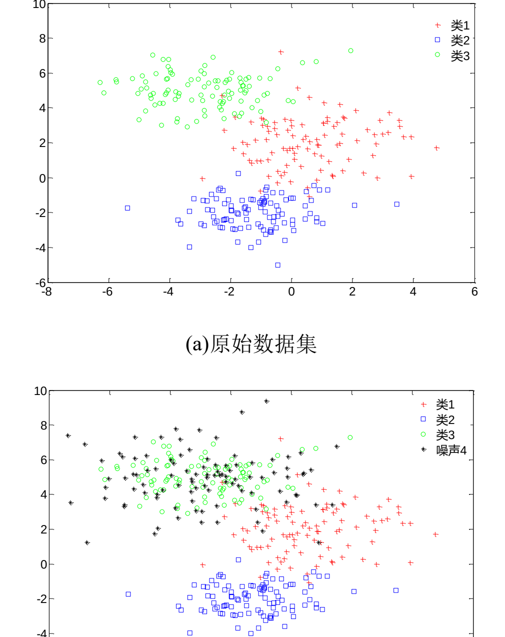
$C_{ij}=\begin{cases}u_{ij}^2,u_{ij}\ge\dfrac{1}{c}\\0,u_{ij}<\dfrac{1}{c}\end{cases}$，是第 i 类和第 j 类样本间的紧致度。当所有样本属于某个类的隶属度较高时，紧致度越大，聚类结果越紧致。总体离散度定义如下：

$$S=\max_{i=1,j=1i\neq j}^{c}S_{ij} \tag{16}$$

其中 $S_{ij}=\min(u_{ik},u_{jk}),k=1,2,\cdots,n$，是第 i 类和第 j 类之间的离散度。两个类别之间划分较清晰时，样本属于某个类的隶属度必定大于其他值，因此，S_{ij} 越小聚类结果越清晰。利用总体离散度衡量最不明确的两个类之间的离散度，当总体离散度越小，表明类之间的差别越大。

本节使用的人工数据集参照文献[16]方法生成，由 2 维欧几里德空间中服从高斯分布的 4 类数据组成，每 1 类均包含 100 个数据。第 1 类中心为(1,2),协方差矩阵为 $\begin{bmatrix}2&0.2\\0.2&2\end{bmatrix}$；第 2 类中心为(-1,-2),协方差矩阵为 $\begin{bmatrix}2&0\\0&1\end{bmatrix}$；第 3 类中心为(-3,5),协方差矩阵为 $\begin{bmatrix}3&0\\0&1\end{bmatrix}$，如图 1(a)所示。第 4 类为置于数据集尾部的噪声样本，中心为(-3,5),协方差矩阵为 $\begin{bmatrix}4&0\\0&3\end{bmatrix}$，如图 1(b)所示。

图 2 给出了 FCM 算法[2]、REFCM 算法[16]和本文 CEFCM 算法对图 1(b)数据集进行聚类所得的划分矩阵结果。可见，三种算法所得的划分矩阵都存在数据样本的某一类隶属度值较高、其余两类隶属度值较低，以及后噪声样本隶属度较低的情况，符合理论上数据样本和噪声样本的隶属度分布各有特点的结论。进一步可以看出，图 2（c）中数据样本和噪声样本隶属度各自分布的特点更加明显，数据样本对各自中心的隶属度值更高，大部分接近 1，而噪声样本的隶属度值均更低，表明本文 CEFCM 算法的数据划分更加合理，噪声对聚类过程的影响明显减小。同时，图 2（c）中数据和噪声的隶属度各自分布的特点越明显，那么它们的差异性越突出，越有利于筛选隶属度值均更低的样本作为噪声，则噪声的识别率更高。



(a)原始数据集

(b)含噪数据集

图 1　人工数据集样本数据

表 1 给出 FCM 算法、REFCM 算法和本文 CEFCM 算法的性能比较。可见，本文算法的紧致度更高，离散度较小，表明本文 CEFCM 算法在样本含噪的情况所获得聚类结果更好，样本划分更清晰，明显具有抗噪强、鲁棒性好的特点。

表 1　3 种聚类算法的性能比较

指标	FCM	REFCM	CEFCM
C	210.3311	88.0310	287.3124
S	0.4465	0.7418	0.3175

3.2　加噪 IRIS 数据集的测试

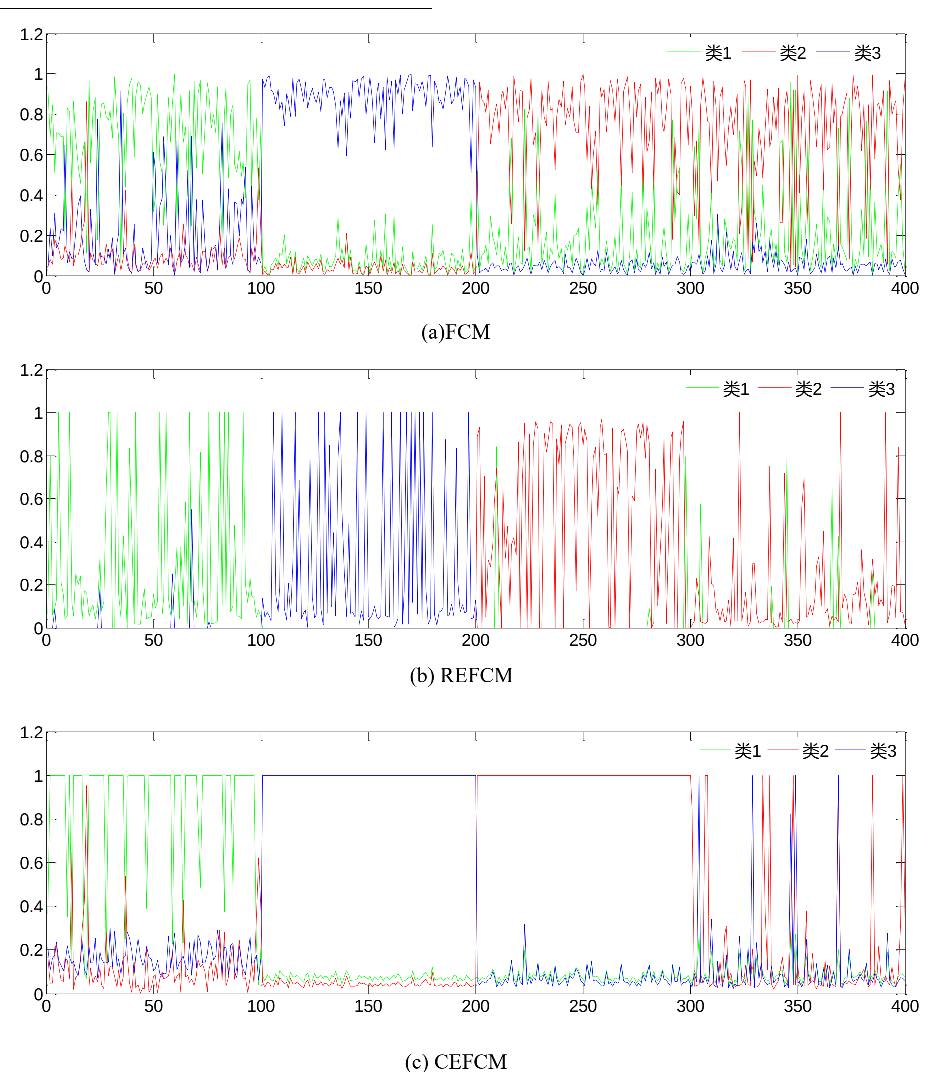
本节采用著名的 IRIS 实际数据集作为测试数据。IRIS 数据集由四维空间中 150 个样本点组成，分为三类，每类有 50 个样本。文献[22]给出了该测试数据的实际类中心位置分别为：p1=(5.00 3.42 1.46 0.24), p2= (5.93 2.77 4.26 1.32), p3= (6.58 2.97 5.55 2.02).

由表 2 可知，在噪声干扰下，FCM、PFCM 和 WFCM 算法的错分数和中心偏差较大，表明算法对噪声敏感，鲁棒性较差，IPCM 和文献[8]方法能够同时具有较好的错分数和聚类中心，表明这两种算法鲁棒性较强，本文方法的错分数最小，正确率最高，中心偏差略低于文献[8]方法，表明本文方法具有较少的错分数和较正确的聚类中心，聚类准确度高。

表 2 各种算法在加噪 IRIS 数据集上的运行结果

算法	FCM	PFCM	IPCM	WFCM	文献[8]方法	本文方法

错分数	50	36	12	50	15	9
中心偏差	1.54	0.44	0.11	1.54	0.22	0.20



(a)FCM

(b) REFCM

(c) CEFCM

图 2　3 种聚类算法的划分矩阵比较

4　结束语

针对 FCM 算法对噪声敏感的问题，本文提出一种基于模糊交叉熵的模糊聚类算法。实验结果表明该算法对噪声鲁棒性强，具有较少的错分数和较正确的聚类中心，聚类结果优于 FCM、PFCM 等现有算法，并且在聚类的同时还可以有效地进行噪声识别。本文算法还需要进一步研究，如参数 λ 的选取、样本特征提取等将是下一步的研究内容。

参考文献：

[1] Ruspini E H. A new approach to clustering [J]. Information & Control, 1969, 15 (1): 22-32.

[2] Bezdek J C. Pattern recognition with fuzzy objective function algorithms [M]. New York: Plenum Press, 1981.

[3] Krishnapuram R, Keller J. A possibilistic approach to clustering [J]. IEEE Trans on Fuzzy Systems, 1993, 1 (2): 98-110.

[4] Zhang Jiangshe, Leung Yiu wing. Improved possibilistic C-means clustering algorithms [J]. IEEE Trans on Fuzzy Systems, 2004, 12 (2): 209-217.

[5] Pal N R, Pal K, Keller J M, *et al.* A possibilistic fuzzy C-means clustering algorithm [J]. IEEE Trans on Fuzzy Systems, 2005, 13 (4): 517-530.

[6] Askari S, Montazerin N, Zarandi M H F. Generalized possibilistic fuzzy C-means with novel cluster validity indices for clustering noisy data [J]. Applied Soft Computing, 2017, 53: 262-283.

[7] 高新波，李洁，姬红兵．基于加权模糊 c 均值聚类与统计检验指导的多阈值图像自动分割算法 [J]．电子学报，2004, 32 (4): 661-664. (Gao Xinbo, Li Jie, Ji Hongbing. A multi-threshold image segmentation algorithm based on weighting fuzzy C-means clustering and statistical test [J]. Acta Electronica Sinica, 2004, 32 (4): 661-664.)

[8] 刘兵，士雄，周勇，等．基于样本加权的可能性模糊聚类算法 [J]．电子学报，2012, 40 (2): 371-375. (Liu Bing, Xia Shixiong, Zhou Yong, *et al.* A sample-weighted possibilistic fuzzy clustering algorithm [J]. Acta Electronica Sinica, 2012, 40 (2): 371-375)

[9] 王丽娜，王建东，李涛，等．集成粗糙集和阴影集的簇特征加权模糊聚类算法 [J]．系统工程与电子技术，2013, 35 (8): 1769-1776. (Wang Lina, Wang Jiandong, Li Tao, *et al.* Cluster’s feature weighting fuzzy clustering

algorithm integrating rough sets and shadowed sets [J]. System Engineering and Electronics, 2013, 35 (8): 1769-1776.)

[10] 赵雪梅, 李玉, 赵泉华. 结合马尔可夫高斯模型的双邻域模糊聚类分割算法 [J]. 计算机辅助设计与图形学学报, 2016, 28 (4): 615-623. (Zhao Xuemei, Li Yu, Zhao Quanhua. A fuzzy clustering image segmentation algorithm with double neighborhood system combined with Markov Gaussian model [J] , Journal of Computer-Aided Design & Computer graphics, 2016, 28 (4): 615-623.)

[11] Zhu Lin; Chung Fulai; Wang Shitong. Generalized fuzzy C-means clustering algorithm with improved fuzzy partitions [J]. IEEE Trans on Systems Man & Cybernetics Part B Cybernetics, 2009, 39 (3): 578-91.

[12] 蒋亦樟, 邓赵红, 王骏, 等. 基于知识利用的迁移学习一般化增强模糊划分聚类算法 [J]. 模式识别与人工智能, 2013, 26 (10): 975-984. (Jiang Yizhang, Deng Zhaohong, Wang Jun, *et al.* Transfer generalized fuzzy c-means clustering algorithm with improved fuzzy partitions by leveraging knowledge [J]. Pattern Recognition and Artificial Intelligence, 2013, 26 (10): 975-984)

[13] Li Ruiping, Mukaidono, M. A Maximum-entropy approach to fuzzy clustering [C]. //Proc of the 4th IEEE International Conference on Fuzzy Systems, Piscataway, NJ: IEEE Press, 1995: 2227-2232.

[14] Wang Shitong, Chung Fulai, Deng Zhaohong, *et al.* Robust maximum entropy clustering algorithm with its labeling for outliers [J]. Soft Computing, 2006, 10 (7): 555-563.

[15] Chaira T. A novel intuitionistic fuzzy C means clustering algorithm and its application to medical images [J]. Applied Soft Computing, 2011, 11 (2): 1711-1717.

[16] Zarinbal M, Zarandi M H F, Turksen I B. Relative entropy fuzzy c-means clustering [J]. Information Sciences, 2014, 260 (1): 74-97.

[17] Li Kai, Cao Zhe. A Fuzzy clustering algorithm with generalized entropy based on neural network [J]. Acta Electronica Sinica, 2016, 44 (8): 1881-1886.

[18] Askari S, Montazerin N, Zarandi M H F, *et al.* Generalized entropy based possibilistic fuzzy C-means for clustering noisy data and its convergence proof [J]. Neurocomputing, 2017, 219 (5): 186-202.

[19] Kullback S. Information Theory and Statistics [M]. New York: John Wiley, 1959

[20] 龙敏, 周铁军. Lambert W 函数性质及其应用 [J]. 衡阳师范学院学报, 2011, 32 (6): 38-40. (Long min, Zhou Tiejun. A survey of properties and application of the Lambert W function [J]. Journal of Hengyang Normal University, 2011, 32 (6): 38-40.)

[21] 朴尚哲, 超木日力格, 于剑. 模糊 C-均值算法的聚类有效性评价 [J]. 模式识别与人工智能, 2015, 28 (5): 452-461. (Piao Shangzhe, Chaomurilige, Yu Jian. Cluster validity indexes for FCM clustering algorithm [J]. Pattern Recognition & Artificial Intelligence, 2015, 28 (5): 452-461.)

[22] Hathaway R J, Bezdek J C. Nerf c-means: Non-Euclidean relational fuzzy clustering [J]. Pattern Recognition, 1994, 27 (3): 429-437.